

## INTISARI

### **PEMODELAN TOPIK UNTUK MEDIA SOSIAL MENGGUNAKAN *BITERM TOPIC MODEL***

oleh

Maria Gaetana Agnesi Tois  
16/406182/PA/17635

Bermedia sosial merupakan aktivitas berinternet yang paling digemari oleh pengguna internet Indonesia. Salah satu media sosial yang paling sering digunakan adalah *Twitter*. Pesatnya pertumbuhan media sosial tersebut menghasilkan informasi berbasis teks yang semakin besar. Dengan banyaknya data berbasis teks, metode analisis teks menyediakan cara yang efektif untuk menggali informasi penting dari data tersebut. Salah satu metode analisis teks yang dapat digunakan adalah pemodelan topik. Pemodelan topik merupakan analisis teks yang bermanfaat dalam pemodelan data tekstual, dengan tujuan menemukan topik yang tersembunyi di dalamnya. Pemodelan topik yang akan dibahas dan digunakan untuk menganalisis data pada skripsi ini adalah *Biterm Topic Model* (BTM). BTM merupakan pengembangan dari *Latent Dirichlet Allocation* (LDA) yang secara efektif dapat memodelkan topik pada teks pendek seperti *Twitter*, berdasarkan pada kumpulan *biterm* dari keseluruhan dokumen. Estimasi parameter yang digunakan dalam model adalah metode Bayesian. Perhitungan estimasi dari distribusi posterior menggunakan algoritma *Gibbs sampling*. Penerapan BTM pada skripsi ini dilakukan pada data tekstual berupa postingan dari tiga akun berita di *Twitter* yaitu @detikcom, @CNNIndonesia, dan @BBCIndonesia untuk mengetahui berita apa yang dominan dibicarakan akun-akun berita tersebut dalam periode tertentu. Hasil pemodelan topik ini berupa proporsi topik pada korpus, probabilitas kata pada setiap topik, dan proporsi topik dalam dokumen/*tweets*.

Kata kunci: analisis teks, pemodelan topik, *Biterm Topic Model*, *Gibbs sampling*, *Twitter*.

## **ABSTRACT**

### **TOPIC MODELS FOR SOCIAL MEDIA USING BITERM TOPIC MODEL**

by

Maria Gaetana Agnesi Tois  
16/406182/PA/17635

Social media is the most popular internet activity among Indonesian internet users. One of the most used social media is Twitter. The rapid growth of social media produces text-based information that is always increasing in amount. With the large amount of text-based data, text analytics provide an effective way to dig useful information from that. One of the text analytic methods that can be used is topic modeling. Topic model is a useful text analysis in modeling textual data, in order to find the hidden topics in it. Topic models that will be discussed and used to analyze data in this thesis is Biterm Topic Model (BTM). BTM is an extension of Latent Dirichlet Allocation (LDA) which can effectively model topics on short texts such as Twitter, based on collections of biterm on the entire documents. Parameter estimation used in this model is the Bayesian method. The estimated calculation of the posterior distribution using Gibbs sampling algorithm. In this paper, BTM probability model is applied for textual data from three news accounts in Twitter that is @detikcom, @CNNIndonesia, and @BBCIndonesia in order to know what dominant topics were being talked by those accounts in a given period. The outcome of this topic model is proportion of topic in corpus, probability of word in each topic, and proportion of topic in document/tweets.

**Keywords:** text analysis, topic models, Biterm Topic Model, Gibbs sampling, Twitter.